method extends a recently proposed procedure by Ohsmann [1] to an order recursive solution. This method does not involve any direct matrix inversion and solves the initial estimate by incorporating a lattice algorithm. We also derived a more accurate expression for the convergence factor which indicates the convergence performance of the iteration (4). The second method is a simplified iterative procedure for situations where the input is known to be white or nonwhite with a correlation coefficient of value $\alpha < 1/3$. This method does not require any matrix inversion and is computationally efficient. Simulation results indicate satisfactory performance of both methods. Results have also been presented for identification in the presence of noise.

Appendix

From (16) we have the matrix $Q = (I_{M+1} - D^{-1}R_M)$. Let us consider that the input $x(k)$ has been obtained as a single pole process $x(k) = \alpha x(k-1) + w(k)$, where $w(k)$ is a zero mean white Gaussian process with unity variance and α is the correlation coefficient. In this derivation, we assume that $0 \le \alpha < 1$ without any loss of generality. We shall also consider that the resulting autocorrelation matrix R_M is symmetric and Toeplitz. Therefore Q has zero entries along the principal diagonal. The characteristic equation $\Delta_m = |Q - \lambda I_{M+1}| = 0$, can be written in the following recursive form [7]:

$$\Delta_m(\lambda) = -\left[2\alpha^2 - \lambda(1+\alpha^2)\right]\Delta_{m-1}(\lambda) - \alpha^2(1+\lambda)^2\Delta_{m-2}(\lambda), \quad m = 1, 2, \cdots, M \tag{A1}$$

where $\Delta_{-1}(\lambda) = 1$, by definition, and $\Delta_0(\lambda) = -\lambda$. After a few manipulations we can rewrite (A1) as

$$\Delta_m(\lambda) = -\left[\alpha(1+\lambda)\right]^m\left[\frac{\lambda\sin(m+1)x}{\sin x} + \frac{\alpha(1+\lambda)\sin mx}{\sin x}\right]. \tag{A2}$$

Equation (A2) then yields $M+1$ distinct eigenvalues given by [7]

$$\lambda_m^M = -2\,\frac{\alpha^2 - \alpha\cos\left(\frac{(m+1)\pi}{M+2}\right)}{1 - 2\alpha\cos\left(\frac{(m+1)\pi}{M+2}\right) + \alpha^2}, \quad m = 0, 1, 2, \cdots, M. \tag{A3}$$

The maximum eigenvalue corresponds to the case $m = 0$ in (A3) and is given by

$$\lambda_{max} = -2\,\frac{\alpha^2 - \alpha\cos\left(\frac{\pi}{M+2}\right)}{1 - 2\alpha\cos\left(\frac{\pi}{M+2}\right) + \alpha^2} \approx \frac{2\alpha}{1-\alpha} \tag{A4}$$

where we let M be large such that $\cos(\pi/M+2) \approx 1$ in (A4). The convergence condition requires that $|\lambda_{max}| < 1$ which yields the bound $\alpha < 1/3$.

References

[1] M. Ohsmann, "An iterative method for the identification of MA systems," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-36, pp. 106-109, no. 1, Jan. 1988.

[2] S. L. Marple, "Efficient least squares FIR system identification," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-29, no. 1, pp. 62-73, Feb. 1981.

[3] S. Haykin, *Adaptive Filter Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1986.

[4] B. Friedlander, "Lattice filters for adaptive filtering," *Proc. IEEE*, vol. 70, no. 8, pp. 829-867, Aug. 1982.

[5] R. S. Varga, *Matrix Iterative Analysis*. Englewood Cliffs, NJ: Prentice-Hall, 1962.

[6] H. Cramér, *Mathematical Methods of Statistics*. Princeton, NJ: Princeton University Press, 1946, pp. 341-363.

[7] U. Grenander and G. Szegö, *Teoplitz Forms and Their Applications*. Berkeley, CA: University of California Press, 1958, pp. 62-80.

[8] L. R. Rabiner, R. E. Chrochiere, and J. B. Allen, "FIR system identification in the presence of noise and with band-limited inputs," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-26, no. 4, pp. 319-333, Aug. 1978.

Weighted Averaging of a Set of Noisy Images for Maximum Signal-to-Noise Ratio

MICHAEL UNSER AND MURRAY EDEN

***Abstract*—This paper considers the problem of estimating a signal from a weighted average of N registered noisy observations. A set of optimal weighting coefficients is determined by maximizing a signal-to-noise ratio criterion. This solution can be computed by first standardizing each observation with respect to its first and second moments and then evaluating the first eigenvector of the corresponding $N \times N$ inner-product matrix. The resulting average is shown to be proportional to the first basis vector of the Karhunen-Loève transform provided that the data has been standardized in an appropriate fashion. The low sensitivity of this approach to the presence of outliers is illustrated by using real electron micrographs of ostensibly identical virus particles.**

I. Introduction

Given a series of noisy observations of an unknown signal, the simplest and often most efficient noise reduction technique is averaging [1, p. 434]. An interesting application of this technique is correlation averaging, which is now used routinely for improving signal-to-noise ratios of electron micrographs of quasi-periodic arrays, or sets of images of ostensibly identical free standing particles [2]-[4]. In this approach, single unit cells or individual particle images are brought into translational and rotational alignment [3]-[5], analyzed using multivariate statistical methods to identify subsets of similar particles and eliminate outliers [6], [7], and finally averaged. The quality of the restored signal is typically assessed in terms of estimated signal-to-noise ratios [4], [8].

An important step in this whole procedure—although not usually emphasized—is the preprocessing where the individual observations are normalized to compensate for varying acquisition parameters. This is usually achieved by simple linear rescaling of the intensity distributions. In most applications, the weighting coefficients are chosen to standardize the range (min/max normalization) or the first and second moments (moment normalization) of the intensity distribution of each observation [9].

In effect, during this analysis the signal is estimated from a weighted average of the initial registered noisy observations. Obviously, there are many ways of selecting the weighting coefficients

Manuscript received October 4, 1988; revised August 2, 1989.

M. Unser is with the Biomedical Engineering and Instrumentation Branch, the National Institutes of Health, Bethesda, MD 20892, and with INSERM, Unit 138, F-94010 Créteil Cedex, France.

M. Eden is with the Biomedical Engineering and Instrumentation Branch, the National Institutes of Health, Bethesda, MD 20892.

IEEE Log Number 9034435.

other than the two forementioned preprocessing techniques. In this paper, we consider these weighting schemes in terms of their performance for signal estimation. In particular, we derive the optimal weighting coefficients that maximize a global signal-to-noise ratio criterion. We then discuss the properties of this solution which is different from a simple standardization and turns out to be closely related to the Karhunen-Loève transform. Finally, we illustrate the use of this technique using simulated and real electron micrograph data, and compare its performance with simpler normalization procedures.

II. The Optimal Solution

The following method applies to the processing of a set of N multidimensional registered noisy observations of a given signal. For notational simplicity, these measurements are represented as vectors in an M-dimensional feature space. This formulation requires the choice of a one-to-one mapping between the discrete coordinate system in the canonical representation (for example, the raster scan of an image) and the components of these feature vectors.

A. Statement of the Problem

Given a set of N noisy measurements, $X = \{x_i, i = 1, \cdots, N\}$, represented as M-dimensional vectors, we are seeking coefficients $\{w_i, i = 1, \cdots, N\}$ and $\{b_i, i = 1, \cdots, N\}$, such that the linearly rescaled vectors

$$z_i = w_i[x_i - b_i\mathbf{1}]; \quad i = 1, \cdots, N \tag{1}$$

define a set $Z = \{z_i, i = 1, \cdots, N\}$ that has a maximum signal-to-noise ratio. In this expression, $\mathbf{1}$ is an M-dimensional vector of all 1's. The signal is estimated from the ensemble average

$$\bar{z} = [\bar{z}_1 \cdots \bar{z}_M]^T = \frac{1}{N}\sum_{i=1}^{N} z_i = \frac{1}{N}\sum_{i=1}^{N} w_i[x_i - b_i\mathbf{1}]. \tag{2}$$

The mean signal value (or dc component) is given by:

$$\bar{\bar{z}} = \frac{1}{M}\sum_{k=1}^{M} \bar{z}_k. \tag{3}$$

The quadratic signal-to-noise ratio (SNR) is defined by the series of equations

$$\text{SNR} \triangleq N\frac{E_{zs}}{E_{zn}} \tag{4}$$

$$E_{zs} \triangleq \sum_{k=1}^{M} [\bar{z}_k - \bar{\bar{z}}]^2 = \bar{z}^T\bar{z} - M\bar{\bar{z}}^2 \tag{5}$$

$$E_{zn} \triangleq \frac{1}{N}\sum_{k=1}^{M}\sum_{i=1}^{N} [z_{ik} - \bar{z}_k]^2 = \frac{1}{N}\sum_{i=1}^{N} [z_i - \bar{z}]^T[z_i - \bar{z}]$$
$$= \frac{1}{N}\sum_{i=1}^{N} \|z_i - \bar{z}\|^2. \tag{6}$$

In other words, E_{zs} and E_{zn} are the rescaled signal and residual noise energies, respectively. The factor N in (4) accounts for the fact that the noise is expected to be reduced by N in the final average. The operator "T" denotes the transpose of a matrix and $\|.\|^2$ is the square norm of a vector. Note that the dc signal component is not taken into account in the definition of the signal energy.

B. Optimal Additive Coefficients

Fortunately, the determination of the optimal additive and multiplicative coefficients in (1) can be carried out independently. This is a consequence of the following theorem.

Theorem 1: For any given set of weights $\{w_i, i = 1, \cdots, N\}$, the b_i's in (1) that maximize the SNR defined by (4)-(6), are given by

$$b_i = \bar{x}_i = \frac{1}{M}\sum_{k=1}^{M} x_{ik}; \quad i = 1, \cdots, N \tag{7}$$

and correspond to the average signal value of each vector.

Proof: Lets us define the centered variables

$$u_i = w_i[x_i - \bar{x}_i\mathbf{1}]; \quad i = 1, \cdots, N \tag{8}$$

which are associated to the average signal $\bar{u}$. It is simple to show that

$$\bar{u} = \frac{1}{M}\mathbf{1}^T\bar{u} = 0$$

and to express the estimated signal $\bar{z}$ as

$$\bar{u} = \bar{u} + \bar{z}\mathbf{1}. \tag{9}$$

By substituting this expression in (5), the signal energy may be rewritten as

$$E_{zs} = \|\bar{z} - \bar{z}\mathbf{1}\|^2 = \|\bar{u}\|^2 \tag{10}$$

and is independent of the value of the b_i's. Similarly, by using (9) and the fact that $z_i = u_i + w_i[\bar{x}_i - b_i]\mathbf{1}$, the noise energy is equal to

$$E_{zn} = \frac{1}{N}\sum_{i=1}^{N} \|u_i - \bar{u} + (w_i[\bar{x}_i - b_i] - \bar{z})\mathbf{1}\|^2.$$

Since the vectors $u_1, \cdots, u_N$, and $\bar{u}$ are all orthogonal to $\mathbf{1}$, i.e., $u_i^T\mathbf{1} = M\bar{u}_i = 0$, this expression is further decomposed as

$$E_{zn} = \frac{1}{N}\sum_{i=1}^{N} \|u_i - \bar{u}\|^2 + \frac{M}{N}\sum_{i=1}^{N} (w_i[\bar{x}_i - b_i] - \bar{z})^2 = E_{un} + E_0. \tag{11}$$

Clearly, E_{un} is independent of the b_i's. Therefore, for any given weights $\{w_i, i = 1, \cdots, N\}$, the SNR is minimum when E_0 is minimum. It is now straightforward to show that the lowest possible value of E_0, namely $E_0 = 0$, is reached when the b_i's are selected according to (7), which also results in $\bar{z} = 0$. Obviously, we can also add any fixed constant to the expression given by (7).

C. Optimal Multiplicative Coefficients

We now proceed to determine the optimum w_i's by replacing the b_i's in (1) by their optimal values: $\bar{x}_i$. We define the $N \times N$ centered inner product matrix

$$\boldsymbol{R}_{xx} = [r_{ij}] \quad \text{with} \quad r_{ij} = x_i^T x_j - M\bar{x}_i\bar{x}_j \tag{12}$$

and the corresponding diagonal matrix

$$\boldsymbol{D}_{xx} = [d_{ij}] \quad \text{with} \quad \begin{cases} d_{ii} = r_{ii} \\ d_{ij} = 0, \quad (i \neq j). \end{cases} \tag{13}$$

The main result of the paper is the following.

Theorem 2: For additive weights in (1) given by (7), the vector of multiplicative coefficients $\boldsymbol{w} = [w_1 \cdots w_N]^T$ that maximizes the SNR defined by (4)-(6) is the first generalized eigenvector of the characteristic equation:

$$\boldsymbol{R}_{xx}\boldsymbol{w} = \beta\boldsymbol{D}_{xx}\boldsymbol{w}. \tag{14}$$

In other words, $\boldsymbol{w}$ is an element of the null space of $(\boldsymbol{R}_{xx} - \beta\boldsymbol{D}_{xx})$, with β the largest value such that there is a nonzero nul space.

Proof: By using (10), we show that the rescaled signal energy is given by

$$E_{zs} = \frac{1}{N^2}\boldsymbol{w}^T\boldsymbol{R}_{xx}\boldsymbol{w}. \tag{15}$$

Similarly, using the centered variables defined by (8), the total energy is equal to

$$E_{zt} = \frac{1}{N}\sum_{i=1}^{N} u_i^T u_i = \frac{1}{N}\boldsymbol{w}^T\boldsymbol{D}_{xx}\boldsymbol{w} \tag{16}$$

and can be decomposed as

$$E_{zt} = \frac{1}{N}\sum_{i=1}^{N} \|u_i\|^2 = \frac{1}{N}\sum_{i=1}^{N} \|u_i - \bar{u}\|^2 + \|\bar{u}\|^2 = E_{zs} + E_{un}. \tag{17}$$

Consequently, maximizing (E_{ss}/E_{nn}) is equivalent to maximizing (E_{ss}/E_{yy}). In addition, the SNR is independent of any global scaling factor so that we can fix the total signal energy to some arbitrary value:

$$E_{yy} = \frac{1}{N} w^T D_{xx} w = \text{const} \tag{18}$$

and simply maximize $E_{ss} = w^T R_{xx} w$ under this constraint. The solution to this optimization problem is found by defining the auxiliary function

$$L(w, \beta) = w^T R_{xx} w - \beta[w^T D_{xx} w - \text{const}] \tag{19}$$

where β is a Lagrange multiplier, and by setting its partial derivatives with respect to w to zero, which yields (14).

The first eigenvector w_1 of (14) is the one that corresponds to the largest eigenvalue:

$$\beta_1 = \max\left\{(w^T R_{xx} w)/(w^T D_{xx} w)\right\}. \tag{20}$$

By decomposing the denominator of this expression into its signal and noise energy components, we find that the maximum signal-to-noise ratio is given by

$$\max\{\text{SNR}\} = \frac{\beta_1}{N - \beta_1}. \tag{21}$$

D. Relationship With the Karhunen-Loève Transform

The process of finding the optimal weights is simplified when D_{xx} is proportional to the identity matrix. This is precisely the case for the normalized variables:

$$y_i = \alpha \frac{x_i - \bar{x}_i \mathbf{1}}{\|x_i - \bar{x}_i \mathbf{1}\|} \tag{22}$$

where α is an arbitrary multiplicative factor, and $\bar{x}_i$ the mean vector value defined by (7). In this particular case, we show that (14) is equivalent to

$$R_{yy} v = \lambda v \tag{23}$$

where v denotes the vector of weights to be applied to the y_i's. The optimal weights are given by the first eigenvector v_1 of the normalized inner product matrix $R_{yy} = Y^T Y$, where $Y = [y_1 \cdots y_N]$ is the $M \times N$ normalized data matrix. In addition, we can show that the optimally weighted average $\bar{z}$ can be directly determined from the Karhunen-Loève (or principal component) representation of the normalized data. More specifically, we would like to emphasize the following properties.

Property 1: The projection of the data matrix on the first eigenvector of R_{yy} is proportional to $\bar{z}$.

Proof:

$$Yv_1 = \sum_{i=1}^{N} v_{1i} y_i = N\bar{z}.$$

Property 2: $\bar{z}$ is proportional to the first eigenvector (or eigensignal) of the $M \times M$ scatter matrix $S_{yy} = YY^T$.

Proof: Using (23), we have that

$$Y^T Y v_1 = \lambda_1 v_1$$

which we multiply on the left by Y

$$Y(Y^T Y)v_1 = \lambda_1 Y v_1$$

$$(YY^T)(Yv_1) = \lambda_1 (Yv_1).$$

By using Property 1, this last expression is also equivalent to

$$(YY^T)\bar{z} = \lambda_1 \bar{z}.$$

Property 3: The nonzero eigenvalues $\{\lambda_i, i = 1, \cdots, N\}$ of $R_{yy} = Y^T Y$ (or $S_{yy} = YY^T$) are directly proportional to the eigenvalues of (14) and the optimum SNR is given by

$$\max\{\text{SNR}\} = \frac{\lambda_1}{\sum_{i=2}^{N} \lambda_i}. \tag{24}$$

Proof: Equation (22) may be rewritten as $y_i = \alpha D_{xx}^{-1/2}[x_i - \bar{x}_i \mathbf{1}]$, which directly implies that

$$R_{yy} = \alpha^2 D_{xx}^{-1/2} R_{xx} D_{xx}^{-1/2}.$$

Substitution of this expression in (23) and left multiplication by $(\alpha^{-2} D_{xx}^{1/2})$ yields

$$R_{xx} D_{xx}^{-1/2} v = (\lambda/\alpha^2) D_{xx}^{1/2} v = (\lambda/\alpha^2) D_{xx}[D_{xx}^{-1/2} v].$$

If we now define $w = D_{xx}^{-1/2} v$, we find that

$$R_{xx} w = (\lambda/\alpha^2) D_{xx} w$$

which is equivalent to (14) provided that $\beta = \lambda/\alpha^2$. We then substitute this value in (21) and note that $\text{tr}\{R_{yy}\} = \alpha^2 N = \sum_{i=1}^{N} \lambda_i$, which finally yields (24).

E. Comments

The maximum signal-to-noise ratio scaling method (MSNRS) was derived by maximizing a global SNR criterion that is entirely specified by the data to be analyzed. This performance criterion, which is only an approximation of the true signal-to-noise ratio, has been used previously to assess the efficiency of correlation averaging techniques [4], [8]. An important point is that the present method makes no assumptions about the underlying signal model and, in this sense, is nonparametric.

From the point of view of statistical estimation theory, a more satisfactory signal estimate would be the one with minimum mean square error (MMSE), which is briefly reviewed in Appendix A. The major difference between MSNRS and MMSE averaging is that the latter is parametric and requires the knowledge of the noise statistics and the scaling coefficients for each image. The MMSE estimation technique is therefore not directly applicable in practice, unless the parameters a and C_{xx} which appear in (A6) are estimated based on some initial approximation of the signal. Despite this difference in philosophy, the equations defining either of these solutions ((14) and (A9), respectively) are structurally very similar, as emphasized in Appendix A.

III. Experimental Results

A set of negatively stained capsomer images of herpes simplex virus (HSV) [10] has been used to illustrate the application of our algorithm. This data was obtained using high resolution electron microscopy. The original micrographs were digitized with a Perkin-Elmer scanning microdensitometer with an effective sampling step of 0.3 nm. Individual capsomer subimages were extracted and brought into rotational and translational alignment using correlation techniques [2]–[5]. Fig. 1 displays the subset of 30 such registered subimages that were used for our analysis. After comparison of the various weighting schemes for the analysis of this particular data set, the extracted signal was used as a test object to further investigate the performance of these algorithms using various levels of Gaussian noise.

A. Analysis of Experimental Electron Micrographs

The original data set was linearly rescaled using (1) with weighting coefficients determined according to the three following procedures:

1) Minimum/maximum normalization (MMN): The minimum and maximum gray level values are constant across all images.

2) Moment normalization (MN): The mean gray level value and energy are constant across all images which implies that the various feature vectors have a constant norm.

3) Maximum SNR scaling (MSNRS).

The resulting averages ($N = 30$) are shown in Fig. 2 with their corresponding signal-to-noise ratios estimated using (4)–(6). Al-

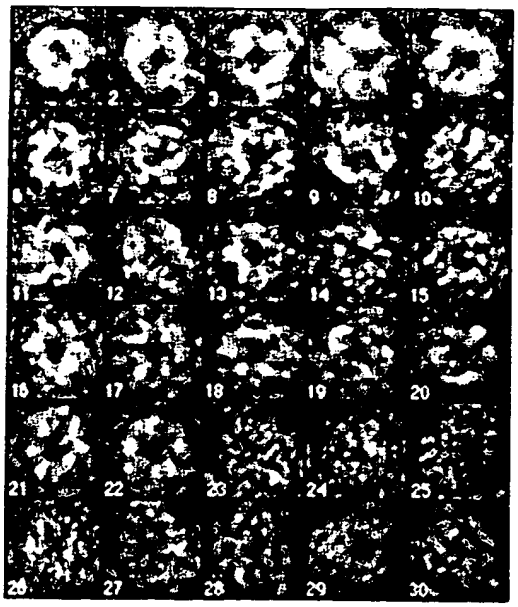

Fig. 1. Set of 30 experimental 50 × 50 capsomer images of herpes simplex virus type II.

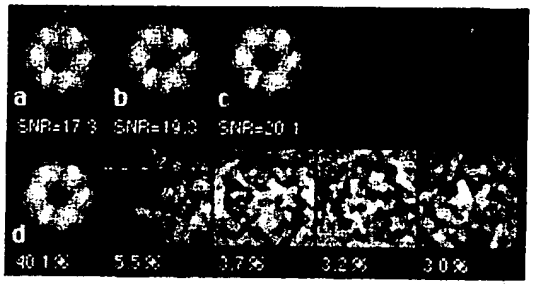

Fig. 2. Comparison of different signal estimation techniques. (a) Average obtained using range normalization. (b) average obtained using moment normalization. (c) maximum signal-to-noise ratio average. (d) 5 first eigen-images of the KLT for standardized images with their corresponding relative energy contributions (eigenvalues).

though these averages are very similar visually, there are differences in the values of the signal-to-noise ratio. The worst results are obtained using the min/max normalization. This is not very surprising since the minimum and maximum values of an image are known to be quite sensitive to the presence of noise. The moment normalization improves the SNR by 11.5%. The MSNRS provides a further 4% increase which is consistent with our theory. We also computed the Karhunen-Loève transform in the case of the images standardized according to the second procedure. The corresponding eigenimages are shown in Fig. 2(d). The first eigenimage is identical to that of Fig. 2(c) and is therefore also very similar to the other averages. We can also verify that the value of the maximum SNR predicted by (24) is identical to the value associated with the optimal average in Fig. 2(c).

B. Outlier Identification

The main advantage of the MSNRS is that it decreases the sensitivity of the analysis to the presence of outliers. To illustrate this effect, we analyzed our data using the odd men out (OMO) algorithm [7]. This procedure uses a criterion of mutual statistical consistency in order to rank the data from most to least reliable. Outliers are then identified by imposing an acceptability threshold on this ordered list. Fig. 3 represents the optimal multiplicative weights as determined by (22) and (23) as a function of the rank produced by the OMO algorithm. The weights consistently decrease as a function of reliability. The Spearman rank correlation is as good as $\rho = -0.977$ indicating that both methods are in excellent agreement. The last six images, which are identified as numbers 14, 24, 28, 27, 23, 25 in Fig. 1, are distinguishable from the bulk of the data in that they are given substantially lower weights. These images probably correspond to outliers, which is fully consistent with the outcome of the OMO algorithm.

The conventional approach to signal extraction using correlation-averaging techniques is to identify outliers using multivariate statistical analyses and to exclude those images from the final average [6], [7], [11]. For this particular example, the OMO algorithm rejects precisely the 6 images identified above [7]. The signal estimate obtained by averaging the remaining 24 images has an SNR of 19.57. This value is extremely close to the one obtained by averaging the full set of 30 images using MSNRS, although the two approaches are based on very different principles.

C. Simulation Experiments

To compare the performance of these algorithms in terms of a fully defined model system, we used the average HSV capsomer image in Fig. 2(c) and contaminated it with increasing levels of noise. We performed two independent experiments using white Gaussian and uniformly distributed noise. For this purpose, we generated a sequence of 20 test images with a quadratic signal-to-noise ratio $\alpha = 1/k^2$, where k is the rank of the corresponding image. The first 5 of these images in the first experiment are displayed in Fig. 4. The averages obtained by using no scaling at all, MMN, MN, and MSNRS, respectively, are shown in Fig. 5. The corresponding measures of the SNR obtained using (3)-(6) are given in Table I. For comparison, we also computed the true signal-to-noise ratio (α) of these averages using the noncorrupted signal (Fig. 2(b)) as reference. Clearly, the best results are obtained for the MSNRS which tends to give the highest weights to the less noisy measurements. The corresponding value of the true SNR ($\alpha = 1.257$ and 1.29) are not too different from the theoretical optimum (minimum mean square error estimate) that could, in principle, be obtained by weighting the images by the inverse of their true variance ($\alpha = 1.5962$, as computed from (A8) in Appendix A). Otherwise, for all other methods, the quality of the signal estimate is substantially below that of even the least noisy image of the set. All scaling algorithms performs essentially the same for Gaussian and uniform noise, with the exception of MMN which is slightly less performant for the latter case.

For this particular example, it is interesting to note that the SNR is a rather optimistic estimate of the true signal-to-noise ratio, particularly for those methods that perform very poorly. This follows from the fact that a badly estimated signal tends to lie much closer on average to the majority of noisy measurements than does the uncorrupted signal. As a consequence, the residual noise variance is usually underestimated leading to an SNR value that can be substantially above the true value. Not surprisingly, the more realistic SNR estimate is obtained for the better performing method, namely the MSNRS.

D. Discussion

In the first example, the averages obtained using either scaling method are very similar perceptually. However, our quantitative analysis shows that the result obtained using the MMN is the less favorable one. The performance of the MN is generally quite satisfactory and the decrement in performance with respect to our optimal approach is only a few percent. We have also performed this evaluation using different sets of micrographs such as the one studied in [4] and [12], and have obtained quite comparable results.

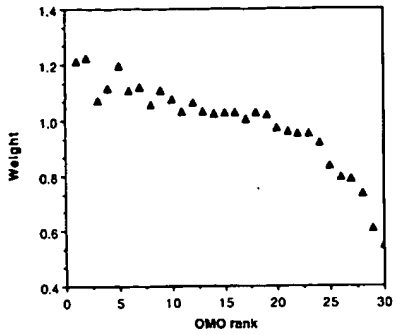

Fig. 3. Maximum signal-to-noise ratio weights for standardized experimental capsomer images in relation to the rank obtained with the OMO algorithm.

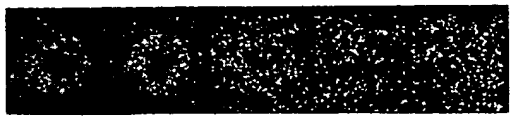

Fig. 4. First 5 images of a total set of 20 HSV capsomers overlaid with increasing amounts of computer-generated Gaussian noise. Signal-to-noise ratios: $\alpha_k = 1/k^2$, $k = 1, \cdots, 20$.

Fig. 5. Comparison of different signal estimation techniques for a generated data set with increasing levels of Gaussian noise. (a) Initial image average, (b) average obtained using range normalization, (c) average obtained using moment normalization, (d) maximum signal-to-noise ratio average.

TABLE I
PERFORMANCE ASSESSMENT OF VARIOUS SCALING METHODS FOR THE AVERAGE OF 20 TEST IMAGES WITH INCREASING LEVELS OF GAUSSIAN AND UNIFORM NOISE

Scaling method	SNR criterion (Gauss, Uniform)	True signal-to-noise ratio (α) (Gauss, Uniform)
none	(1.16, 1.21)	(0.14, 0.15)
minimum/maximum normalization (MMN)	(1.53, 1.48)	(0.53, 0.45)
moment normalization (MN)	(1.57, 1.57)	(0.55, 0.56)
maximum SNR scaling (MSNRS)	(1.96, 1.94)	(1.26, 1.29)

In analyzing these data sets, we observed that the main advantage of the MSNRS lies in its ability to handle outliers by decreasing their relative contribution to the final average.

To further differentiate between these approaches, we have considered the simulated example of a signal corrupted by various levels of Gaussian noise. For this extreme case, the MSNRS alone resulted in satisfactory signal extraction, owing to the fact that it weights the images approximately in proportion to their signal-to-noise ratio. The performance of the other weighting schemes was less than adequate, with a signal estimate that was even noisier than the best image of the set. An important point brought out by this example is that the relative increase of the true signal-to-noise ratio may be substantially greater than the corresponding improvement of the SNR criterion (cf. Table I). As pointed out earlier, this is due to the fact that the noise component tends to be underestimated for poor signal estimates. This result suggests that for real micrographs such as those shown in Fig. 1, the improvement of MSNRS over the other schemes may be even greater than that which may be inferred from the relative values of the SNR criterion. This assertion, however, remains speculative since there is no way to measure the true signal-to-noise ratio for real noisy data sets.

Based on these observations, it appears that the simple MN is adequate for analyzing reasonably homogeneous data sets; that is, apart from some constant scaling factors, the individual observations are approximately identically distributed. The MSNRS is the most useful in less ideal cases when the data at hand is likely to contain outliers or when the individual observations are subject to different levels of perturbation.

In contrast with the OMO algorithm [7] which operates on a simple acceptance/rejection basis, the MSNRS treats outliers in a graded manner and is closer to the principle underlying fuzzy set theory. In this sense, the weight given to a particular image expresses our confidence that it belongs to the main group. We also remark that unlike the OMO algorithm our present derivation makes no assumptions about the underlying signal model.

According to property 2, when the data has been properly normalized the first eigenvector of the KLT is proportional to the maximum signal-to-noise ratio signal estimate. This result provides a mathematical explanation for the great similarity that can usually be observed between the first basis vector of the Karhunen-Loève transform (or principal components) and the average of the data. In fact, the difference between those two vectors is significant only when the MSNRS clearly outperforms the MN. Conversely, for very nearly homogeneous data sets (e.g. images with a common mean and identically distributed noise), the first eigen-coordinates should be essentially a constant across all images.

Finally, we remark that in many applications of KLT or principal components analysis the main group effect is intentionally factored out by subtracting the mean vector and performing the analysis on centered variables. Our study provides mathematical evidence to justify the physical judgments that underly the use of this factor.

IV. Summary and Conclusion

In this paper, we have considered the problem of estimating a signal from a weighted average of N noisy observations or feature vectors. In this approach, the intensity distribution of each observation is scaled linearly to compensate for varying acquisition parameters. The optimal weights have been derived by maximizing a global signal-to-noise ratio criterion. The determination of the MSNRS solution involves the standardization of the individual observations by subtraction of the mean intensity value and division by the square root of the residual energy. The optimal weights are then obtained from the components of the first eigenvector of the corresponding inner-product matrix.

An interesting conceptual result is the link that has been established between the first basis vector of the Karhunen-Loève transform for normalized feature vectors and the average vector: both of these vectors are estimates of the underlying signal with the former being the one with maximal signal-to-noise ratio.

The MSNRS provides an upper bound on performance for comparison with other simple scaling methods. We have shown that its most interesting feature lies in its capacity to adjust to the presence of outliers. It is also particularly suited for analyzing data sets of uneven quality where the initial signal-to-noise ratio can vary substantially from one observation to another. Other than that, for reasonably homogeneous data sets, a simple normalization (mean subtraction and constant norm) of the feature vectors prior to averaging leads to a signal estimate that is only slightly suboptimal. The MSNRS is therefore best suited for what we have referred to as less

than ideal cases. A good example of application is high resolution electron microscopy where it is difficult to collect images that are fundamentally alike due to differences in the levels of noise, variation of viewing geometry, nonuniformity of staining, or other structural perturbations.

Appendix A
Minimum Mean Square Error Averaging

In this Appendix, we derive the multiplicative weights for minimum mean square error averaging, assuming that the noise statistics as well as scaling factors associated with each observation are known. For simplicity, we consider the case of one-dimensional feature vectors. The results are easily extended to higher dimensions by replacing the μ^2 by the norm of the signal vector and the between-measurement covariance matrix C_{xx} by the average of these matrices across all vector components.

Given a set of N noisy measurements $\{x_i\}$ such that

$$E\{x_i\} = a_i\mu \tag{A1}$$

where $\{a_i\}$ is a set of known constants, the problem is to determine an estimation of the signal parameter μ from the weighted sum:

$$\hat{\mu} = \sum_{i=1}^{N} w_i x_i \tag{A2}$$

with the requirement that

$$E\{\hat{\mu}\} = \mu \tag{A3}$$

such that the mean square error $\epsilon^2 = E\{(\hat{\mu} - \mu)^2\}$ is minimized. We note that this solution will also maximize the signal-to-noise ratio defined as $\alpha = \mu^2/\epsilon^2$. The first condition implies that

$$\sum_{i=1}^{N} w_i a_i = \mathbf{w}^T\mathbf{a} = 1 \tag{A4}$$

where $\mathbf{w} = [w_1 \cdots w_N]^T$ and $\mathbf{a} = [a_1 \cdots a_N]^T$. It is straightforward to show that the residual error is given by

$$\epsilon^2 = \mathbf{w}^T\mathbf{C}_{xx}\mathbf{w} \tag{A5}$$

where $\mathbf{C}_{xx} = [c_{ij}]$ is the between-measurement covariance matrix whose entries c_{ij} are defined as:

$$c_{ij} = E\{(x_i - E\{x_i\})(x_j - E\{x_j\})\}.$$

Taking the partial derivatives of (A5) and (A4) with respect to $\mathbf{w}$, we find an expression that is consistent with standard results on linear unbiased minimum variance estimation [13, ch. 5]:

$$\mathbf{w} = \lambda\mathbf{C}_{xx}^{-1}\mathbf{a} \tag{A6}$$

where λ is a Lagrange multiplier which should be chosen to satisfy (A4). If the measurements are mutually uncorrelated, we get the well-known result that the observations should be given a weight that is inversely proportional to their variance:

$$w_i = (a_i\lambda)/c_{ii}. \tag{A7}$$

It is relatively easy to show that the corresponding maximum SNR is given by:

$$\alpha^* = \sum_{i=1}^{N} \alpha_i \tag{A8}$$

where $\alpha_i = (a_i\mu)/c_{ii}$ is the SNR associated to the random variable x_i. We now rewrite (A6) to bring out the similarities with (14). The simplest form of (A6) is

$$\mathbf{C}_{xx}\mathbf{w} = \lambda\mathbf{a}$$

which, by noticing that $\mathbf{a}\mathbf{a}^T\mathbf{w}$ is always in the direction of $\mathbf{a}$, is also equivalent to

$$\mu^2(\mathbf{a}\mathbf{a}^T)\mathbf{w} = \beta[\mathbf{C}_{xx} + \mu^2(\mathbf{a}\mathbf{a}^T)]\mathbf{w}. \tag{A9}$$

Equation (A9) is now in a form that is comparable with (14). In particular, the matrix $\mu^2(\mathbf{a}\mathbf{a}^T)$, which is of rank 1, is the parametric equivalent of the matrix $\mathbf{R}_{xx}$ in (14). For instance, the expected signal energy is given by:

$$E\{\hat{\mu}\}^2 = \mathbf{w}^T[\mu^2(\mathbf{a}\mathbf{a}^T)]\mathbf{w} = (\mu\mathbf{w}^T\mathbf{a})^2 \tag{A10}$$

which is similar to (15). We also note that $[\mathbf{C}_{xx} + \mu^2(\mathbf{a}\mathbf{a}^T)]$ is the parametric equivalent of $\mathbf{D}_{xx}$ since the total energy of the estimated signal is given by

$$E\{\hat{\mu}^2\} = \epsilon^2 + E\{\hat{\mu}\}^2 = \mathbf{w}^T[\mathbf{C}_{xx} + \mu^2(\mathbf{a}\mathbf{a}^T)]\mathbf{w} \tag{A11}$$

which is similar to (19).

Acknowledgment

The authors thank Dr. A. C. Steven and Dr. B. L. Trus for helpful discussions and for making available the electron micrographs.

References

[1] W. K. Pratt, *Digital Image Processing*. New York: Wiley, 1978.
[2] W. O. Saxton, *Computer Techniques for Image Processing in Electron Microscopy*. New York: Academic, 1978.
[3] J. Frank, "Averaging of low-exposure electron micrographs of nonperiodic objects," *Ultramicroscopy*, vol. 1, pp. 159-162, 1975.
[4] J. Frank, A. Verschoor, and M. Boublik, "Computer averaging of electron micrographs of 40S ribosomal subunits," *Science*, vol. 214, pp. 1353-1355, 1981.
[5] J. P. Secilla, N. Garcia, and J. L. Carrascosa, "Template location in noisy pictures," *Signal Processing*, vol. 14, pp. 347-361, 1988.
[6] M. Van Heel, "Multivariate statistical classification of noisy images (randomly oriented biological macromolecules)," *Ultramicroscopy*, vol. 13, pp. 165-184, 1984.
[7] M. Unser, A. C. Steven, and B. L. Trus, "Odd men out: A quantitative objective procedure for identifying anomalous members of a set of noisy imnages of ostensibly identical specimens," *Ultramicroscopy*, vol. 19, pp. 337-347, 1986.
[8] M. Unser, B. L. Trus, and A. C. Steven, "A new resolution criterion based on spectral signal-to-noise ratio," *Ultramicroscopy*, vol. 23, pp. 39-52, 1987.
[9] M. Unser, B. L. Trus, and A. C. Steven, "Normalization procedures and factorial representations for classification of correlation-aligned images: A comparative study," *Ultramicroscopy*, vol. 30, pp. 299-310, 1989.
[10] A. C. Steven, C. R. Roberts, J. Hay, M. E. Bisher, T. Pun, and B. L. Trus, "Hexavalent capsomers of herpes simplex virus type 2: Symmetry, shape, dimensions, and oligomeric status," *J. Virology*, pp. 578-584, Feb. 1986.
[11] M. Van Heel and J. Frank, "Use of multivariate statistics in analyzing images of biological macromolecules," *Ultramicroscopy*, vol. 6, pp. 187-194, 1981.
[12] A. C. Steven, B. L. Trus, J. V. Maizel, M. Unser, D. A. D. Parry, J. S. Wall, J. F. Hainfeld, and F. W. Studier, "The molecular structure of a viral receptor-recognition protein—the GP17 tail-fiber of bacteriophage T7," *J. Mol. Biol.*, vol. 200, pp. 351-365, 1988.
[13] F. C. Schweppe, *Uncertain Dynamic Systems*. Englewood Cliffs, NJ: Prentice-Hall, 1973.